

# Conversational based method for tweet contextualization

Rami Belkaroui<sup>1</sup>  · Rim Faiz<sup>2</sup>

Received: 29 April 2016 / Accepted: 23 December 2016 / Published online: 9 January 2017  
© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** Bound to 140 characters, tweets are short and ambiguous by nature. It can be hard for a user without any kind of context to effectively understand what the tweet is about. Due to this restriction, it is, therefore, necessary to know the tweet's context to make it easily understandable to a reader. In this paper, we treat the problem of tweet contextualization. We propose a specific method allowing to automatically contextualize tweets using information coming from social user interactions. Contrary to classical contextualization methods that only consider text information which is insufficient, since text information on Twitter is very sparse, we combine different types of signals (social, temporal, textual). Our experimental results validate the benefits of our approach and confirm that generated contexts contain relevant information with given tweet.

**Keywords** Tweet contextualization · Social influence · Twitter conversations · Social context

## 1 Introduction

Since the recent few years, microblogging has become a very popular form of communication that attracts more and more users due to the ease and the speed of information sharing. Daily, people are posting millions of updates under the form

of short text messages on microblogging services, such as Twitter<sup>1</sup> and Facebook<sup>2</sup>. The size of these status updates is limited by a maximum number of characters. This limitation causes the use of a special vocabulary that is not usually used, noisy and full of new words [8]. Indeed, the purpose is to share the maximum amount of information in few characters [18]. It may thus be difficult to understand the meaning of a short text message without knowing the general context of its realization. This constraint problem is, for example, a frequent case on Twitter microblogging platform.

As Twitter gains popularity, a lot of messages are generated daily, allowing users to communicate with each other and share different variety of information. Moreover, Twitter's data are examined to measure public sentiment [12], earthquake warning [24], follow political activity and news. For both end-users and data analysts, it is a hard task to plow through millions of tweets which contain a lot of noise and redundancy. Furthermore, since a tweet is short and without sufficient contextual information, it is often difficult to understand the related information. All these difficulties impede users from effective understanding or consuming information, which can make users less engaged in using Twitter.

Twitter is both a microblogging service and a conversational environment that enables people to interact, engage in daily chatter and join conversations. Conversations are key element in such service. Almost a quarter of Twitter users hold conversations with other users through this platform [14] and huge percentage of Twitter posts are conversational [21]. The public aspect of conversations makes Twitter one of the available publicity available resources of naturally occurring conversations. The huge volume of conversations produced everyday makes it an interesting information source. Thus,

✉ Rami Belkaroui  
rami.belkaroui@gmail.com

Rim Faiz  
rim.faiz@ihc.rnu.tn

<sup>1</sup> LARODEC, ISG Tunis, University of Tunis, Bardo, Tunis, Tunisia

<sup>2</sup> LARODEC, IHEC Carthage, University of Carthage, Carthage Presidency, Carthage, Tunisia

<sup>1</sup> <http://www.twitter.com>.

<sup>2</sup> <http://www.facebook.com>.

exploiting Twitter conversations to provide context for a given tweet is the main contribution of this paper.

In the previous proposed contextualization approaches, Wikipedia are the most used knowledge source to extract a bunch of relevant sentences that present some additional information about the tweet's context. Interestingly, sometimes after news events such as earthquakes or other natural disasters, Wikipedia information is not directly available. Even, after a few hours, the available content is often imprecise. For example, regarding the Charlie Hebdo attack, there were no articles on Wikipedia describing the topic #jesuis-charlie. Indeed, the first article that explains this event was available 7h after the terrorist attack. At the same time, these events describe a scenario where users urgently need information, especially if they are directly concerned by the event. So, unexpected news events such as earthquakes represent information access problem where the approaches using Wikipedia to contextualize a tweet decline.

In this paper, we extend our model presented in [5] and propose a new method based on social conversations, which helps users to get more context information when using Twitter. Our contributions are manifold. We explore the social influence as well as several social features for context generation. In order to measure influence, we propose a user-tweet model allowing us to capture user- and tweet-based characteristics that we can consider as influence markers. We also consider multiple types of signals such as social signals (hashtags, URLs), temporal signals and text-based signals, which can be potentially useful to improve tweet contextualization task.

The rest of this paper is organized as follows: we begin by describing some related works. In Sect. 3 a detailed description of our method for tweet contextualization is presented. Our experimental results are exposed in Sect. 4. Finally, we conclude and introduce some future work in Sect. 5.

## 2 Related work

In this section, we discuss several areas of works related to our proposed tweet contextualization method such as information retrieval, automatic summarization, entity linking and Twitter content categorization.

### 2.1 Information retrieval and automatic summarization

In recent years, various studies focused on the problem of short message contextualization. Almost all of the proposed approaches have combined information retrieval and automatic summarization techniques. In [22], the authors took advantage of a larger use of hashtags and used them to enhance the retrieval of relevant Wikipedia articles. Moreover, the proposed approach in [7] described a hybrid tweet

contextualization system using information retrieval and automatic summarization. They used nutch architecture, TF-IDF based sentence ranking and sentence extracting techniques for automatic summarization. In [2], the authors have simply treated contextualization as a passage retrieval task. They used textual tweet content as a query to retrieve paragraphs or sentences from the Wikipedia corpus. In [11], the authors used a method that allows to automatically contextualize tweets by using information coming from Wikipedia. They treat the problem of tweet contextualization as an automatic summarization task, where the text to resume is composed of Wikipedia articles that discuss the various pieces of information appearing in a tweet.

### 2.2 Entity linking and semantic modeling

Some researchers have focused on exploiting semantic aspect and named entities mentioned in tweets in order to solve tweet contextualization problem. In [6], the authors proposed an algorithm to detect the subjects of microblogs based on named-entity recognition. Then, a search engine is used to discover more information about these entities to define the tweet context. In the same way, [17] proposed an entity-based profiling approach, which aims to discover the topics of interest for Twitter users by examining entities they mention in their tweets and leveraging Wikipedia as a knowledge source. In [16], the authors proposed a machine learning-based approach using n-gram features and tweet features to identify concepts semantically related to a tweet. Similarly, [1] proposed an approach based on matching tweets to news articles, followed by semantic enrichment based on news article's content.

### 2.3 Twitter content analysis and categorization

A number of studies investigated Twitter content analysis and categorization. In [18], the authors used Latent Dirichlet Analysis (LDA) to obtain a tweet representation in a thematic space. This representation allows finding a set of latent topics covered by the tweet, which should help to better understand a given tweet. In [10], the authors proposed a Twitter content classification framework as a tool for personal, professional, commercial and phatic communications happen in real-world application based on 16 existing Twitter studies and a grounded theory analysis of a personal Twitter history. In [9], a classification scheme for tweet categorization based on textual content and its underlying structural information are presented. In [28] Schultz et al. proposed a hashtag-based categorization approach in order to test the applicability of several machine learning techniques in conjunction with relevant feature selection algorithms for the text categorization problem. In the same way, [13] introduced a Wikipedia-based classification technique. They classified

tweets by mapping message into their most similar Wikipedia pages and calculating semantic distances between messages based on distances between their closest Wikipedia pages.

### 3 Proposed method for tweet contextualization

In this paper, we propose a specific method, depicted in Fig. 1, based on Twitter conversations to provide more context information for a given tweet. This method is performed on the following two steps:

1. Retrieving relevant Twitter conversations containing information related to initial tweet (tweet extension).
2. Extracting the most salient tweets from retrieved conversations to build context.

In the following section, we detail each step of our proposed method.

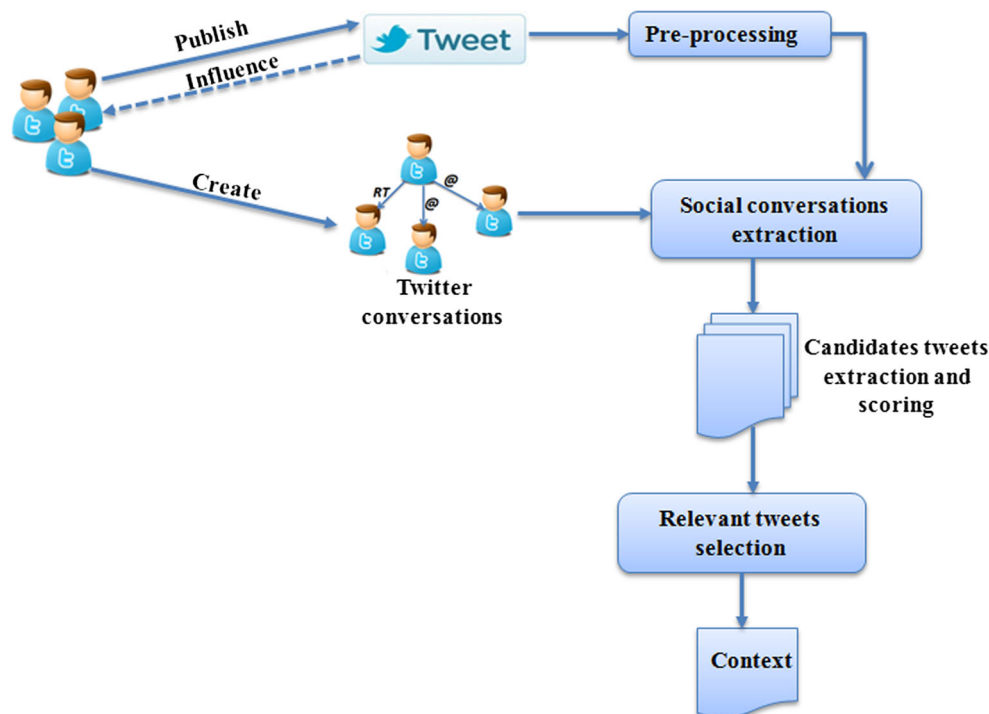
#### 3.1 Basic concepts definitions

- Tweet: we represent a tweet as bag of words. We removed all the stop words (based on the standard INQUERY stop list). The final representation is a clean tweet without stop words or useless words. We used  $t$  as a symbol for tweet object.
- User: symbol  $u$  is used for user object.

- Initial tweet representation: we represent an initial (ambiguous) tweet  $t_{in}$  as a bag of hashtags. Formally, we have  $t_{in_i} = \{h_1, \dots, h_j\}$
- tweet contextualization task: the idea is to expand a collection of  $n$  initial (ambiguous) tweets  $S_{in} = \{t_{in_1}, \dots, t_{in_n}\}$  using a collection of  $m$  Twitter conversations  $S_c = \{c_1, \dots, c_m\}$  by providing a context  $C_i$  for each tweet  $t_{in_i} \in S_{in}$ . For given tweet, we retrieve a sub-set  $sub_c$  of relevant conversations from  $S_c$ ; then we select the most relevant tweets from conversations in  $sub_c$ .
- Context representation: the context  $C_i$  of an initial tweet  $t_{in}$  is defined as a set of informative tweets from the  $sub_c$  sub-set.
- Twitter conversation: we define Twitter conversation as a set of tweets posted by users at specific timestamp on the same topic. These tweets can be directly replied to other users by using “@username” or indirectly by retweeting, mention and other possible interactions (favorite).

#### 3.2 Twitter conversations trees analysis

In this part, we conduct further analysis on Twitter conversations trees with respect to temporal growth and depth distributions. Owing to our conversation trees detection system, we collect 5000 Twitter conversations from January 15th to March 30th, 2015. We focus on the collected data set in the following:



**Fig. 1** Overview of our proposed framework

### 1. Temporal growth analysis

Figure 2 presents the temporal growth of the Twitter conversations, where  $y$  axis is the number of tweets and  $x$  axis is the relative temporal distance from the original tweets, measured by hours. Given that Twitter is a real-time service, overall, about 97.87% of replies are generated within the first hour, while an additional 0.98% of replies happen in the second hour, which shows that Twitter can propagate information quite fast and a meaningful context tree can be formed very quickly. Consequently, the temporal growth of the context tree prove the importance of exploiting twitter conversations in our method.

### 2. Depth distributions analysis

The depth of conversation tree is defined as the maximal distance from the root [15, 19]. Figure 3 shows the cumulative distribution of the number of tweets over depth in conversation trees, where  $y$  axis is the percentage of depth distribution and  $x$  axis is distribution of depth levels. Surprisingly, the structures of Twitter context trees are highly skewed, and more than 80% of tweets are at depth 1 (assuming that the depth each tree root is 0). In addition, 10.7% of Twitter conversations have two levels depth. Only 1.53% of Twitter conversations have three levels depth. This distribution means that increasing the reply level decreases the information content of the tweet, while increasing the conversation length increases the information content.

## 3.3 Candidate tweets retrieval from social conversations

In this part, we focus on one crucial step of our proposed method to retrieve conversations that are relevant to the initial tweet. Hopefully, these conversations contain a set of informative tweets that provides enough contextual information to (fully) understand the meaning of a given tweet.

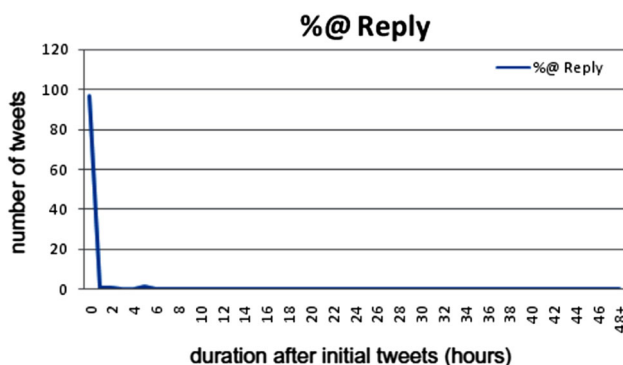


Fig. 2 Number of tweets over time

### 3.3.1 Initial tweet formatting

Hashtags, in tweets, are very important pieces of information, since they are tags that were generated by Twitter users as a way to categorize their messages. Hashtags are used to mark keywords or topics in a tweet. In addition, users can use hashtags to provide implicit tweet context. Furthermore, we view a tweet's hashtag content is a good approximation of its total content [23]. We consider hashtags as the tweet's keywords, because they normally are names or places and it, therefore, seems logical to favor their use in the context of a recovery social conversations related to initial tweet.

For each initial tweet, we applied a formatting process. We removed all the retweet mentions (RT), user mentions (@username), and stop words (based on the standard INQUERY stop list) from the tweets. The final output of the formatting process is a set of hashtags. Furthermore, the final set of hashtags is used as a short keyword query to retrieve conversations. Figure 4 shows an example of our initial tweet formatting process and the final query used to retrieve conversations.

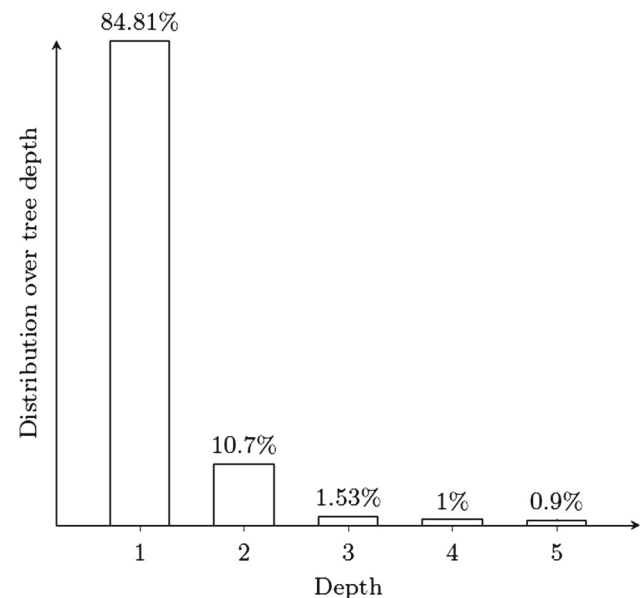


Fig. 3 Distributions over tree depth



Fig. 4 Initial tweet formatting

### 3.3.2 Retrieving Twitter conversations

Given an initial tweet, we select the top 10 relevant conversations from the retrieved collection, as relevant. Then we calculate different characteristics for each of these tweets that allow us to classify and form the context. In the following, we describe, in detail, the candidate tweet selection step.

### 3.4 Candidate tweets selection

We calculated different features for each candidate tweet to further rank them and generate the tweet's context. There are three categories of features:

- **tweet influence:** the importance of a tweet within the conversation where it appears is estimated using social influence.
- **tweet relevance regarding initial text:** we compute the cosine similarity between the candidate tweet and the initial tweet.
- **tweet relevance regarding URL:** we compute the word overlap and the cosine similarity between the candidate tweet and the body content of the linked page, as well as with the title of the Web page.

#### 3.4.1 Social influence generation based on user–tweet interaction model

On Twitter, there are various interactional relationships between users and tweets such as post, reply, mention and retweet, depicted in Fig 5. We profit from these relationships in order to measure tweet influence score and select context candidate tweets. We exploit two types of score for tweet influence measuring the following:

- **tweet influence score:** refers to those features which represent the particular characteristics of tweet.
- **tweet's author influence score:** refers to those features which represent the influence of tweet's author.

#### Tweet influence measuring

The tweet influence is determined by reply, retweet and favorite influence.

- **Reply influence score(t)** The action here is replying. The more replies a tweet receives, the more influential it is. This influence can be measured by the number of replies that the tweet receives. The reply influence is defined as follows:

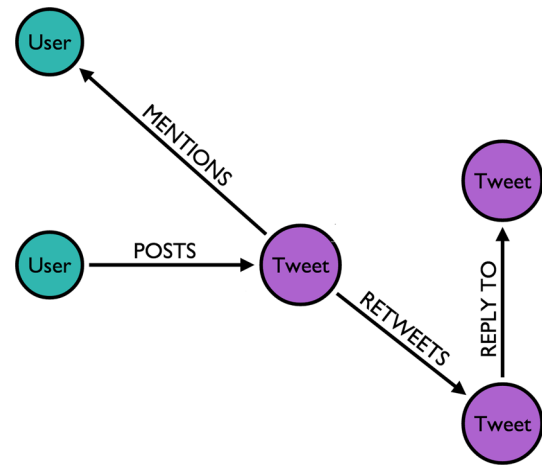


Fig. 5 Relationships between users and tweets

$$\text{Reply\_influence}(t) = \alpha \times \text{number\_reply}(t). \quad (1)$$

$\alpha \in (0, 1]$ . It is adjustable and indicates the weight of reply edge.

- **Retweet influence score(t)** The action here is retweeting. The more frequently user's messages are retweeted by others, the more influential it is. This can also be quantified by the number of retweet. It is defined as follows:

$$\text{Retweet\_influence}(t) = \beta \times \text{number\_retweet}(t). \quad (2)$$

$\beta \in (0, 1]$ . It is adjustable and indicates the weight of retweet edge.

- **Favorite influence score(t)** The action here is favoriting. When a user mark a tweet as favorite, she/he indicates that the tweet's content is useful and relevant. The more favorites a tweet receives, the more influential it is. This influence can be determined by the number of favorite the tweet receives. It is defined as follows:

$$\text{Favorite\_influence}(t) = \gamma \times \text{number\_favorite}(t). \quad (3)$$

$\gamma \in (0, 1]$ . It is adjustable and indicates the weight of favorite edge.

Due to the real-time nature of Twitter, we consider that exploiting temporal aspect can provide valuable information

for tweet contextualization problem. In addition, tweet timestamp plays an important role on tweet influence, i.e., a recent tweet has larger chance to have bigger influences compared to old published tweet. So to cope with, we use Gaussian Kernel [20] to calculate a difference  $\Delta t$  between tweet root time  $t_{root}$  and other tweets times  $t$  within the same conversation, i.e.,  $\Delta t = |t - t_{root}|$ . It is defined as follows:

$$\Gamma(\Delta t) = \exp \left[ \frac{-\Delta t^2}{2\sigma^2} \right] \quad \text{with } \sigma \in \mathbb{R}^+. \quad (4)$$

Finally, the tweet influence score is defined as follows:

$$\begin{aligned} tweet\_influence(t) = & \Gamma(\Delta t) \times Reply\_influence(t) \\ & + Retweet\_influence(t) \\ & + Favorite\_influence(t). \end{aligned} \quad (5)$$

#### Tweet's author influence measuring

On Twitter, celebrities post many messages that obtain an import number of followers because of their real-life influence. In addition, considering only the number of followers cannot represent the real user influence. It has been confirmed that features expressing engaging audience links such as mention relationship are better to represent user influence. Furthermore, in our approach we consider both follow relationship and the mention relationship.

- **Mention influence** measured through the number of mentions containing one's name, it indicates the ability of that user to engage others in a conversation. The mention influence score is defined as follows:

$$Mention\_influence(u) = \delta \times number\_Mention(u). \quad (6)$$

$\delta \in (0, 1]$ . It indicates the weight of mention edge.

- **Follow influence**

A user followed by many other users is likely to be an authoritative user and their posts are also likely to be useful. In addition, the follower number can directly indicate the user audience size. The follow influence score is defined as follows:

$$Follow\_influence(u) = \omega \times number\_follow(u). \quad (7)$$

$\omega \in (0, 1]$ . It indicates the weight of follow edge.

Finally, the tweet's author influence score is defined as follows:

$$\begin{aligned} tweetauthor\_influence(u) = & Mention\_influence(u) \\ & + Follow\_influence(u) \end{aligned} \quad (8)$$

We describe in the Table 1, how we set influence parameters. We define that the weight of  $\alpha$  is bigger than  $\beta$  and  $\gamma$  weights, which means that the users who reply on tweet  $t$  are more interested in it than others who only retweet or favorite it. In addition, when we calculate  $\delta$  and  $\omega$ , we give the same values equals to the threshold 0.5.

### 3.5 Candidate tweets scoring

We assign score to a candidate tweet based on the similarity between different tweets in the whole conversation. Therefore, from each tweet  $t$  in a conversation  $C$ , we derive a vector  $\vec{V} = \{w_1, w_2, \dots, m_i\}$  as a set of words using the vector space model [25].

- Similarity to initial tweet

We used cosine similarity to calculate the similarity between initial tweet vector  $\vec{V}_{in}$  and other tweets vectors  $\vec{V}_t$  within the same conversation. In addition, we aim to measure how much a tweet would be related to initial tweet content.

$$\cosine(\vec{V}_t, \vec{V}_{in}) = \frac{\vec{V}_t \cdot \vec{V}_{in}}{\|\vec{V}_t\| \cdot \|\vec{V}_{in}\|} \quad (9)$$

- Similarity of content

In our approach, we measure how many tweets of the whole conversation  $C$  are similar in content with current tweet  $t_{current}$ . We calculate cosine similarity score for every pair of tweets. The similarity is calculated using lucene similarity function. We denote current tweet modeled as a vector:

$$\cosine(t_{current}, C) = \frac{\sum_{t_{current} \neq \vec{t}} similarity(t_{current}, \vec{t})}{|C| - 1} \quad (10)$$

**Table 1** Parameter weights setting

Parameter	Weight
$\alpha$	0.6
$\beta$	0.2
$\gamma$	0.2
$\delta$	0.5
$\omega$	0.5



- Relevance regarding URLs

When an URL is present in the tweet, we download the page and extract its title as well as the body content. For each candidate tweet  $t$ , we computed the following:

- The word overlap between a candidate tweet  $t$  and the web page title, and between  $t$  and the body content of the web page.
- The cosine similarity between  $t$  and the web page title, and between  $t$  and the body content of the web page.

We measure the final importance score of a candidate tweet as a linear combination of the above features. The weights were learned and are presented in Table 2. After every tweet has been attributed a score, they are ordered and the top-ranked tweets are selected to form context.

## 4 Experiments and results

In this section, we detail our experimental setup, including how we sample tweets, conversations dataset, reference summary and how we evaluate relating individual tweets to their contexts. Thus, our reference collection to evaluate tweet contextualization task contains the following:

- A collection of relevant conversations: these are conversations that are used as a resource to extract components (tweets) of a context corresponding to the contextualization of a tweet.
- A collection of tweets to contextualize: they correspond to a set of ambiguous tweets.
- Reference summary: that will be used to compare their contents to our proposed summary.
- Evaluation measures: tweet contextualization is evaluated on both informativeness and readability; we will use these measures to evaluate our results.

**Table 2** Feature weights

Feature	Name	Weight
c1	Tweet influence	0.6257
c2	Tweetauthor influence	0.533
c3	Cosine initial tweet	0.207
c4	Cosine tweet	0.3128
c5	Overlap text URLs	0.459
c6	Cosine titre URLs	0.025

### 4.1 Ambiguous tweets dataset

The tweets dataset has been collected by monitoring Twitter microblogging system over the period of January–March 2015. In particular, we have built a real tweets collection. Thus, we manually selected 100 tweets, so that

- We have selected only tweets among informative accounts (e.g. @CNN) to avoid purely personal tweets that could not be contextualized.
- We have chosen only tweets containing hashtags; therefore, hashtags can be considered as one of the main topics of the tweet. This type of information, hashtags, will be used in our experiments to improve the queries used to retrieve conversations and, therefore, to improve the generated content.

### 4.2 Twitter conversations dataset

We extract 5000 Twitter conversations from January 15th to March 30th, 2015, using our conversation tree detection system [4] to construct a data set in our work. This conversation tree contains related content to the collection of tweets. We cleaned our collection of conversations by filtering out the conversations that involve less than 3 participants and containing less than 5 tweets. Also, we just keep the conversations that contain over 3 hashtags related to the initial tweet.

### 4.3 Reference summary

To the best of our knowledge, there is no dataset available to evaluate tweet contextualization task. In order to create a reference summary, we conduct a pilot study to construct an editorial dataset generated by a set of assessors that can be useful to evaluate our results. In addition, the assessors selected among students and colleagues of the authors (with backgrounds in computing and social sciences). Thus, for each initial tweet, we only consider the top 10 best conversations and we ask 10 assessors to judge every context tree. In addition, we ask each assessor to read the initial tweet and open any URL inside to have an idea about this tweet. Then, the assessor reads all candidates' tweets and selects 5 to 10 tweets ordered sequentially as a context, which extend the initial tweet by providing additional information about it.

### 4.4 Evaluation metrics

Tweet contextualization is evaluated on both informativeness and readability [26]. Informativeness aims at measuring how well the summary explains tweet or how well the summary helps user to understand the tweet content. On the other hand,

readability aims at measuring how clear and easy is to understand the summary.

- **Informativeness** The objective of this metric is to evaluate relevant tweets selection. The 10 best tweets summary, for each initial tweet, are selected for evaluation. This choice is made based on the score assigned by the automatic system tweet contextualization (high scores). The dissimilarity between a human selected summary (constructed using a pilot study) and the proposed summary (using our method) is given by

$$Dis(T, S) = \sum_{t \in T} (P - 1) \times \left( 1 - \frac{\min(\log(P), \log(Q))}{\max(\log(P), \log(Q))} \right), \quad (11)$$

where  $P = \frac{f_T(t)}{f_T} + 1$  and  $Q = \frac{f_S(t)}{f_S} + 1$ .

$S$  is the set of informative tweets presented in our proposed summary. and  $T$  is the set of terms presented in reference summary. For each term  $t \in T$ ,  $f_T(t)$  represents the frequency of occurrence of  $t$  in reference summary and  $f_S(t)$  its frequency of occurrence in the proposed summary. The More the  $Dis(T, S)$  is low, the more the proposed summary is similar to the reference.  $T$  may take three distinct forms:

- Unigrams made of single lemmas.
- Bigrams made of pairs of consecutive lemmas (in the same sentence).
- Bigrams with 2-gaps as well as the bigram, but can be separated by two lemmas.

Our results in the informativeness evaluation are presented in Table 3.

- **Readability** Readability aims at measuring how clear and easy it is to understand summary. By contrast, readability is evaluated manually and presented in Table 4. Each summary has been evaluated by considering the following parameters [27]:
  - Relevance: judge if the tweet make sense in their context (i.e. after reading the other tweets in the same context). Each assessor had to evaluate relevance with three levels, namely highly relevant (value equal to 2), relevant (value equal to 1) or irrelevant (value equal to 0).
  - Non-redundancy: evaluates the ability of context does not contain too much redundant information, i.e., information that has already been given in a previous tweet. Each assessor had to evaluate redundancy with three levels, namely not redundant (value equal

**Table 3** Table of informativeness results

	Unigrams	Bigrams	Skipgrams
Topic1			
Human summary	0.7263	0.8534	0.9213
Proposed summary	<b>0.7009</b>	<b>0.8165</b>	<b>0.9055</b>
Topic2			
Human summary	0.7932	0.9137	0.9361
Proposed summary	<b>0.7505</b>	<b>0.9008</b>	<b>0.9192</b>
Topic3			
Human summary	0.7786	0.9472	0.9526
Proposed summary	<b>0.7127</b>	<b>0.9138</b>	<b>0.9117</b>

to 2), redundant (value equal to 1) or highly redundant (value equal to 0).

- Soundness: each assessor had to evaluate the anaphora resolution in the context.
- Syntax: each assessor had to evaluate syntax of produced context.

We presented in this paper a method allowing to automatically contextualize tweets. A good context should have good quality but with less redundancy. Informativeness evaluation, presented in Table 3, involves computation of three metrics: the dissimilarity between a human selected summary and the proposed summary for uni-grams, bi-grams, and bi-grams with two allowable gaps in between. Note that dissimilarity being a distance measure implies that a lower value of this metric is indicative of a better result. The obtained informativeness evaluation results explain that our proposed method offers interesting results and ensure that context contain adequate correlating information with the initial tweet. The results of our experiments suggest that the use of hashtags present in tweets helps in retrieving relevant conversations that contain elements providing contextual information. In addition, by examining the influence of different characteristics, we found that user-tweet influence information is very helpful to generate a high-quality context from Twitter conversations. Besides, we found that the set of conversation retrieved using the tweet as query are the best candidates for the generation of contexts. Furthermore, we note that the tweets selection effects the context quality and enhance the informativeness. The contexts are less readable; it may be that they contain some noises which need to be cleaner.

#### 4.5 Comparison using INEX's data

We use the collection dataset of tweet contextualization INEX track 2013 to compare our method with the official results. The output contexts were evaluated according to their



**Table 4** Table of readability results

	Relevance (%)	Non redundancy (%)	Soundness (%)	Syntax (%)	AVG (%)
Topic1					
Human summary	88.65	66.33	65.04	69.22	72.31
Proposed summary	89.72	69.78	70.68	67.37	74.38
Topic2					
Human summary	90.72	65.82	68.24	71.52	74.07
Proposed summary	91.03	67.49	74.52	70.02	75.76
Topic3					
Human summary	90.23	69.06	65.04	67.34	72.91
Proposed summary	90.24	69.72	66.64	62.35	72.23

**Table 5** Comparison of our produced context and the best run in terms of informativeness score at INEX 2013, 2014

	Unigrams	Bigrams	Skipgrams
ref2013	0.705	0.794	0.796
ref2014	0.7528	0.8499	0.8516
Proposed method	0.7709	0.702	0.855

informativeness. For the 2013 edition, the organizers explicitly introduce a larger number of tweets containing hashtags. We indexed the collection with the Indri free search engine by removing keywords present in the INQUERY list.

In Table 5 are reported the performances of our produced context with the two best official results at INEX 2013 and INEX 2014 in terms of informativeness. We note that the results are relatively similar and there is no significant difference between the three approaches. The weak differences observed between the three results are probably due to the relative similarity between models of IR, even if we see that the use of hashtags significantly improves scores. We notice that the results obtained by our method using INEX's data are not the best; also we note that the performance of our method is reduced because the most significant social features used to estimate the relevance of a sentence could not be applied to the INEX's data. In addition, measures of similarity between sentences and tweets are a reliable indicators, while the hashtags here seem to have a random influence. We conclude that our method is very efficient and has very good results with a social dataset of conversations while the results are reduced with INEX's data.

## 5 Conclusion

We explored in this paper the tweet contextualization problem. We extend our model presented in [3] and we proposed a specific method that combined different types of signals from

social user interactions and exploited a set of conversational features, which help users to get more context information when using Twitter. We focused on exploiting multiple types of signals such as social signals, user-tweet influence signals and text-based signals. In our ongoing research, we would deepen our method by gathering multiple data sources such as comments on news articles or on Facebook pages.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Semantic enrichment of twitter posts for user profile construction on the social web. In: Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web: Research and Applications, vol. Part II. ESWC'11, pp. 375–389. Springer, Berlin (2011)
2. Bandyopadhyay, A., Pal, S., Mitra, M., Majumder, P., Ghosh, K.: Passage retrieval for tweet contextualization at INEX 2012. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17–20 (2012)
3. Belkaroui, R., Faiz, R.: Towards events tweet contextualization using social influence model and users conversations. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, WIMS 2015, Larnaca, Cyprus, July 13–15, pp. 3:1–3:9 (2015). doi:[10.1145/2797115.2797134](https://doi.org/10.1145/2797115.2797134)
4. Belkaroui, R., Faiz, R., Elkhilfi, A.: Social users interactions detection based on conversational aspects. In: Barbucha, D., Nguyen, N.T., Batubara, J. (eds.) New Trends in Intelligent Information and Database Systems. Studies in Computational Intelligence, vol. 598, pp. 161–170. Springer International Publishing, Berlin (2015)
5. Belkaroui, R., Faiz, R., Kuntz, P.: User-tweet interaction model and social users interactions for tweet contextualization. In: Computational Collective Intelligence—7th International Conference, ICCCI 2015, Madrid, Spain, September 21–23, 2015. Proceedings, Part I, pp. 144–157 (2015). doi:[10.1007/978-3-319-24069-5\\_14](https://doi.org/10.1007/978-3-319-24069-5_14)
6. Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., Chi, E.H.: Eddi: interactive topic-based browsing of social status streams. In:

- Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. UIST '10, pp. 303–312. ACM, New York (2010)
7. Bhaskar, P., Banerjee, S., Bandyopadhyay, S.: A hybrid tweet contextualization system using IR and summarization. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17–20 (2012)
  8. Choudhury, M., Saraf, R., Jain, V., Mukherjee, A., Sarkar, S., Basu, A.: Investigation and modeling of the structure of texting language. *Int. J. Doc. Anal. Recognit. (IJ DAR)* **10**(3–4), 157–174 (2007)
  9. Cotel, J., Cruz, F., Enríquez, F., Troyano, J.: Tweet categorization by combining content and structural knowledge. *Inf. Fusion* **31**, 54–64 (2016)
  10. Dann, S.: Twitter content classification. *First Monday* **15**(12) (2010)
  11. Deveau, R., Boudin, F.: Contextualisation automatique de tweets à partir de wikipédia. In: CORIA 2013—Conférence en Recherche d'Informations et Applications—10th French Information Retrieval Conference, Neuchâtel, Suisse, April 3–5, pp. 125–140 (2013)
  12. Gaffney, D.: iranelection: quantifying online activism. In: WebSci10: Extending the Frontiers of Society On-Line (2010)
  13. Genc, Y., Sakamoto, Y., Nickerson, J.V.: Discovering context: classifying tweets through a semantic transform based on wikipedia. In: Proceedings of the 6th International Conference on Foundations of Augmented Cognition: Directing the Future of Adaptive Systems. FAC'11, pp. 484–492. Springer, Berlin (2011)
  14. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: Understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. WebKDD/SNA-KDD '07, pp. 56–65. ACM, New York (2007)
  15. Kumar, R., Mahdian, M., McGlohon, M.: Dynamics of conversations. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10, pp. 553–562. ACM, New York (2010). doi:[10.1145/1835804.1835875](https://doi.org/10.1145/1835804.1835875)
  16. Meij, E., Weerkamp, W., de Rijke, M.: Adding semantics to microblog posts. Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. WSDM '12, pp. 563–572. ACM, New York (2012)
  17. Michelson, M., Macskassy, S.A.: Discovering users' topics of interest on twitter: a first look. In: Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data. AND '10, pp. 73–80. ACM, New York (2010)
  18. Morchid, M., Linares, G.: Inex 2012 benchmark a semantic space for tweets contextualization. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF (Online Working Notes/Labs/Workshop) (2012)
  19. Nishi, R., Takaguchi, T., Oka, K., Maehara, T., Toyoda, M., Kawarabayashi, K.i., Masuda, N.: Reply trees in twitter: data analysis and branching process models. *Social Netw. Anal. Min.* **6**(1), 1–13 (2016)
  20. Phillips, J.M., Venkatasubramanian, S.: A gentle introduction to the kernel distance. *Comput. Res. Reposit.* (2011). [arXiv:1103.1625](https://arxiv.org/abs/1103.1625)
  21. Ritter, A., Cherry, C., Dolan, B.: Unsupervised modeling of twitter conversations. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. HLT '10, pp. 172–180. Association for Computational Linguistics, Stroudsburg (2010)
  22. Romain, D., Florian, B.: Effective tweet contextualization with hashtags performance prediction and multi-document summarization. In: Working Notes for CLEF 2013 Conference, Valencia, Spain, September 23–26 (2013)
  23. Rosa, K.D., Shah, R., Lin, B., Gershman, A., Frederking, R.: Topical clustering of tweets. In: SWSM, Proceedings of the ACM SIGIR (2011)
  24. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web. WWW '10, pp. 851–860. ACM, New York (2010)
  25. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. *Communications of the ACM*. **11**, pp. 613–620. ACM, New York (1975)
  26. SanJuan, E., Bellot, P., Moriceau, V., Tannier, X.: Overview of the inex 2010 question answering track (qa@inex). In: Geva, S., Kamps, J., Schenkel, R., Trotman, A. (eds.) Comparative Evaluation of Focused Retrieval. Lecture Notes in Computer Science, vol. 6932, pp. 269–281. Springer, Berlin (2011)
  27. SanJuan, E., Moriceau, V., Tannier, X., Bellot, P., Mothe, J.: Overview of the index 2012 tweet contextualization track. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF (Online Working Notes/Labs/Workshop) (2012)
  28. Schultz, D., Jolly, S.: Automatic Tweet Hashtag Categorization (2010)